

WHAT IS CLAIMED IS:

1. A computer-implemented method comprising:  
receiving a document;  
determining a file type for the document;  
segmenting the document into blocks of text as a  
function of the file type; and  
generating at least one keyword and a summary  
for the document.
2. The computer-implemented method of claim 1  
wherein segmenting includes using outline information  
to segment text in the document.
3. The computer-implemented method of claim 1  
wherein segmenting includes analyzing HTML tags in  
the document and segmenting text in the document  
based on the HTML tags.
4. The computer-implemented method of claim 3  
wherein segmenting further includes using a position  
of text in the html document.
5. The computer-implemented method of claim 1 and  
further comprising providing at least one keyword and  
a summary for each block of text in the document.
6. The computer-implemented method of claim 1 and  
further comprising displaying the at least one  
keyword and summary.

7. The computer-implemented method of claim 1 and further comprising generating at least one keyword and a summary for individual blocks of text within the document.
8. The computer-implemented method of claim 1 and further comprising establishing potential segmentation points based on text in the document.
9. The computer-implemented method of claim 8 and further comprising determining final segmentation points based on similarity of adjacent blocks of text surrounding the potential segmentation points.
10. The computer-implemented method of claim 1 and further comprising converting an audio file to a text document.
11. A computer-implemented method comprising:
  - retrieving a plurality of documents;
  - determining a file type for each of the plurality of documents;
  - segmenting each of the plurality of documents into blocks of text as a function of the file type;
  - determining at least one keyword and a summary for each of the plurality of documents; and
  - providing an output of the at least one keyword and summary for each of the plurality of documents.

12. The computer-implemented method of claim 11 wherein segmenting includes using outline information to segment text in the document.

13. The computer-implemented method of claim 11 wherein segmenting includes analyzing HTML tags in the document and segmenting text in the document based on the HTML tags.

14. The computer-implemented method of claim 13 wherein segmenting further includes using a position of text in the html document.

15. The computer-implemented method of claim 11 and further comprising providing at least one keyword and a summary for each block of text in the document.

16. The computer-implemented method of claim 11 and further comprising displaying the at least one keyword and summary for each document.

17. The computer-implemented method of claim 11 and further comprising generating at least one keyword and a summary for individual blocks of text within each of the plurality of documents.

18. The computer-implemented method of claim 11 and further comprising establishing potential

segmentation points based on text in each of the plurality of documents.

19. The computer-implemented method of claim 18 and further comprising determining final segmentation points based on similarity of adjacent blocks of text surrounding the potential segmentation points.

20. The computer-implemented method of claim 11 and further comprising converting an audio file to a text document.

21. A computer-readable medium, comprising:  
a document retrieval module adapted to retrieve a document from a document source based on a document query request received from a mobile device;  
a document outline parsing module adapted to determine a file type of the document and segment the document into blocks of text based on the file type; and  
a summarization module adapted to generate at least one keyword and a summary for the document.

22. The computer-readable medium of claim 21 wherein the document outline parsing module is further adapted to segment the document using outline information associated with the document.

23. The computer-readable medium of claim 21 wherein the document outline parsing module is further adapted to analyze HTML tags in the document and segment text in the document based on the HTML tags.

24. The computer-readable medium of claim 23 wherein the document outline parsing modules is further adapted to segment the document using a position of text in the html document.

25. The computer-readable medium of claim 21 wherein the summarization module is further adapted to provide at least one keyword and a summary for each block of text in the document.

26. The computer-readable medium of claim 21 and further comprising a module adapted to display the at least one keyword and summary.

27. The computer-readable medium of claim 21 wherein summarization module further is adapted to generate at least one keyword and a summary for individual blocks of text within the document.

28. The computer-readable medium of claim 21 wherein the document outline parsing module is further adapted to establish potential segmentation points based on text in the document.

29. The computer-readable medium of claim 21 wherein the document outline parsing module is further adapted to determine final segmentation points based on similarity of adjacent blocks of text surrounding the potential segmentation points.

30. The computer-readable medium of claim 21 and further comprising a module adapted to convert an audio file to a text document.